

Homework Sheet — Stage 03: Python Fundamentals

Due: Aug 15 (next session)

Submission: Push to your GitHub repo.

Assignment

In the lecture, we learned about core Python data structures, NumPy, pandas, and reusable function design. Now, you will adapt these skills to explore and summarize a dataset.

Step-by-Step Instructions

1. NumPy Operations

- Create an array and perform elementwise operations.
- Compare loop vs vectorized execution.

2. Dataset Loading

- Load provided CSV (`data/starter_data.csv`) using pandas.
- Inspect with `.info()` and `.head()`.

3. Summary Statistics

- Calculate `.describe()` for numeric columns.
- Perform `.groupby()` aggregation by category.

4. Save Outputs

- Save summary stats to `data/processed/summary.csv` or `.json`.
- Bonus: Create and save a basic plot.

5. Reusable Functions

- Write at least one utility function (e.g., `get_summary_stats(df)`).
 - Bonus: Move function to `src/utils.py` and import in notebook.
-

Grading Rubric (100 points)

- **NumPy operations (10 pts)**
 - **Data loading & inspection (10 pts)**
 - **Summary statistics (15 pts)**
 - **Groupby aggregation (15 pts)**
 - **Output saved correctly (10 pts)**
 - **Utility function written (15 pts)**
 - **Code readability & comments (10 pts)**
 - **Bonus tasks (+5 pts)**
-

Example Submission

- `notebooks/hw03_python_fundamentals.ipynb` with all steps.
 - `data/processed/summary.csv` file.
 - Optional: `src/utils.py`.
-